REMARKS

The examiner's stated basis for restriction is that Nagase *et al.*, *DNA Research* 3: 43-53 (1996) ("Nagase II"), "teaches a KIAA0172 gene," which the examiner asserts to be the "technical feature linking groups I – VII" (action at page 3, lines 1 – 3). With this perspective, the examiner concludes that the "technical feature" in question "does not define a contribution over the art" (*id.*, lines 5 – 6).

Applicants would point out, however, that Nagase I is supplemental to another, contemporaneous article by Nagase *et al.*, *DNA Research* 3: 17-24 (1996) ("Nagase I"), which is Exhibit A to this response. Nagase I discloses a cDNA clone, designated "KIAA0172," said to encode a polypeptide of 1307 amino acids (see Table 3 of Nagase I).

The polypeptide thus disclosed actually does not exist in any intact cell, however; this, because Nagase I makes an erroneous assignment of the first codon. Instead, for a correct assignment of the first codon the mRNA cap site had to be determined, via extensive and elaborate experimental investigation, by the present inventors.

Accordingly, the present application confirms the absence from cells of a polypeptide as taught by Nagase I. Likewise incorrect is the GenBank sequence, D79994, that corresponds to the Nagase I disclosure and that the authors deposited. A copy of the D79994 sequence is Exhibit B to this response.

It is apparent, therefore, that the present application is the first to disclose a correct sequence for "a KIAA0172 gene." That correct sequence is applicants' SEQ ID NO: 1, which has 1194 amino acids.

This 1194 amino-acid sequence differs from the Nagase I sequence, which has 1367 amino acids. Furthermore, neither Nagase I nor Nagase II discloses a function for the protein encoded by KIAA0172. Indeed, the protein posited by Nagase I differs so markedly from the protein having the sequence of SEQ ID NO:1 that one could not reasonably predict that the former has the same function as the present application describes for the latter.

In light of the foregoing, it is evident that the examiner's stated basis for restriction is factually untenable. Accordingly, what he has identified as the technical feature linking the invention of groups I – VIII in fact does define a contribution over the prior art represented by Nagase I/II. The respective claim sets of Groups I – VIII should not be divided, therefore, but rather should be examined together.

Respectfully submitted,

Date 10 July 2006

By S. A. Bent

FOLEY & LARDNER LLP
Customer Number: 22428
Telephone: (202) 672-5404
Facsimile: (202) 672-5399

Stephen A. Bent
Attorney for Applicant
Registration No. 29,768

DNA RESEARCH 3, 17–24 (1996)

Prediction of the Coding Sequences of Unidentified Human Genes. V. The Coding Sequences of 40 New Genes (KIAA0161-KIAA0200) Deduced by Analysis of cDNA Clones from Human Cell Line KG-1

Takahiro NAGASE, Naohiko SEKI, Ken-ichi ISHIKAWA, Ayako TANAKA, and Nobuo NOMURA*
Kazusa DNA Research Institute, 1532-3 Yanauchino, Kisarazu, Chiba 292, Japan

(Received 1 February 1996; revised 21 February 1996)

Abstract

As part of our continuing efforts to accumulate information on the coding region of unidentified human genes, we newly determined the sequences of 40 cDNA clones of human cell line KG-1 which correspond to relatively long and nearly full-length transcripts, and predicted the coding sequences of the corresponding genes, named KIAA0161 to 0200. The average size of the cDNA clones analyzed was approximately 5.0 kb. A computer search of the sequences in public databases indicated that the sequences of 20 genes were unrelated to any reported genes, while the remaining 20 genes carried sequences which show some similarities to known genes. Among the genes in the latter category, KIAA0167 contained a Zn-finger motif with significant structural similarity to that of the yeast transcription factor *GCS1*, and KIAA0189 was classified into the RhoGAP gene family. Stretches of typical CAG (Gln) repeats, which were often correlated with genetic disorders, were found in KIAA0181 and KIAA0192. Another novel repeat composed of alternating Arg and Glu was identified in KIAA0182. Northern hybridization analysis demonstrated that 10 genes are expressed in a cell- or tissue-specific manner.

Key words: full-length cDNA sequence; CAG repeat; transcriptional factor; RhoGAP gene family; myeloid cell line KG-1

1. Introduction

In this series of projects involving the accumulation of information on the coding sequences of unidentified human genes, we have been analyzing nearly full-length cDNA clones which were isolated from human immature myeloid cell line KG-1.[1] We already reported the sequences of 160 new cDNA clones and predicted the coding regions of the corresponding genes.[1–4] The average size of these cDNA clones, except for 20 clones (KIAA0101 to 0120), was approximately 4.0 kb. Each clone contained a distinct open reading frame (ORF) in the 5′-moiety, and their avarage size was roughly 1.7 kb, indicating that most of the clones carried relatively long 3′-untranslated regions (3′-UTRs). Although vast amounts of expressed sequence tags (ESTs) obtained by single-run sequencing of cDNA libraries have been accumulated for comprehensive understanding of expression profiles, our preliminary analysis indicated that most of ESTs (GenBank release 92.0, Dec. 1995) fell in the region about 2 kb from the poly(A)-tail of our cDNA sequences. This is probably due to the fact that the cDNA libraries prepared by conventional methods contain a fairly large amount of small clones derived from truncated transcripts. On the basis of computer analysis, biological function has been predicted for at least 40% of the genes that we reported, from which the functional significance of 20 genes is under investigation in collaboration with other laboratories. We keep on sequencing the new cDNA clones, and in this paper, we report the coding sequences of 40 additional genes and their sequence features as well as expression profiles.

Communicated by Mituru Takanami

* To whom correspondence should be addressed. Tel. +81-438-52-3930, Fax. +81-438-52-3931

2. Materials and Methods

The source of cDNA libraries and methods used for selection of cDNA clones, Northern hybridization, sequence analysis, computer analysis of sequences and chromosomal mapping of cDNA clones were described previously.[1]

3. Results and Discussion

3.1. Sequence features of analysed cDNA clones

The cDNA clones carrying inserts longer than 2 kb were randomly selected from the libraries constructed from the middle-sized cDNA class, and both the terminal sequences were analysed to select unidentified clones with poly(A) tails.[1] The clones carrying inserts which were more than 90% of the length of the corresponding transcripts were further selected by Northern hybridization,

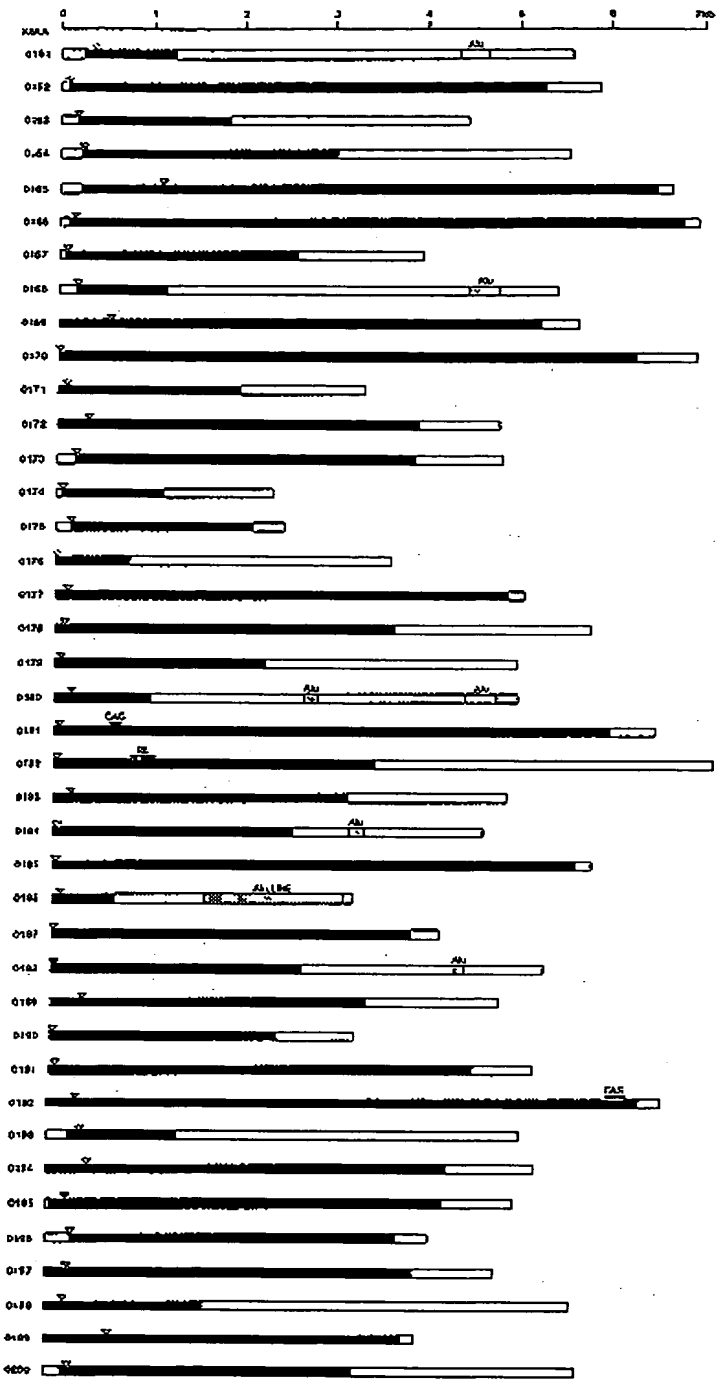


Fig. 1. Physical maps of the 40 cDNA clones analyzed. The horizontal scale represents the cDNA length in kb, and gene numbers are given on the left. Open reading frames (ORFs) within coding regions, untranslated regions and repetitive sequences are indicated by solid, open and dotted boxes, respectively. The positions of the first ATG codon in each ORF are represented by triangles. The names of repetitive sequences are described above the dotted boxes. The solid bars show the positions of the triplet and other repeats listed in Fig. 2. The nucleotide sequence data reported in this paper were deposited in the GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases under the accession numbers shown in Table 3.

Table 1. cDNA clones with similarities to PIR and GenBank/EMBL database files.

Gene no. (KIAA)	Database files	Accession no.[a)]	Identity (%)	Overlap[b)] (amino acid residues)
0162	emb-5 protein	S35241	31.5	1525
0164	DNA-binding protein 5 (H)	S26650	38.4	159
0165	cut1 protein (Sp)	A35694	36.0	236
0167	KIAA0041 (H)	D26069[c)]	46.0	126
	KIAA0050 (H)	D30758[c)]	35.5	166
0170	calphotin (D)	A47283	17.8	841
0171	hypothetical protein L8167.6 (Sc)	S48557	30.8	214
0172	ankyrin 3, long form (H)	A55575	34.5	139
0173	tubulin-tyrosine ligase (P)	A45443	30.4	214
0175	protein kinase p69Eg3 (X)	S52244	62.1	655
0177	poly(ADP-ribose)synthase (C)	JH0581	29.0	303
0178	SMC1 protein (Sc)	A49464	28.2	1241
0179	hypothetical protein D4478 (Sc)	S48776	33.8	130
0185	hypothetical protein YM9959.11C (Sc)	S57596	26.0	1087
0188	SMP2 protein (Sc)	S30911	47.9	240
0189	regulator protein p122 - RhoGAP (R)	S54293	40.1	1039
0190	Ubiquitin-specific proteinase UBP3 (Sc)	B44450	23.9	368
0192	Mopa box protein (M)	A26892	94.6	129
0198	finger protein (clone XlcOF7.1) (X)	S06546	29.0	214
0199	hydroxymethylglutaryl-CoA reductase (Su)	A31898	28.0	168
0200	neurogenic locus mam protein (D)	A36391	18.5	243

[a)] PIR database files are shown except for KIAA0167. [b)] The size of regions which show similarities. [c)] GenBank/EMBL database files.
C, chicken; Ce, *Caenorhabditis elegance*; D, *Drosophila melanogaster*; H, human; M, mouse; P, pig; R, rat; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Su, sea urchin; X, *Xenopus laevis*.

Table 2. cDNA clones with regions that matched motifs in the PROSITE database.

Motifs	Description	Gene number (KIAA)	References
ZINC FINGER C3HC4	Zinc finger, C3HC4 type	0161	13
ATP GTPA	ATP/GTP-binding site motif A (P-loop)	0167, 0178, 0187	14
PROTEIN KINASE ST	Protein kinases	0175	15
ABC TRANSPORTER	ABC transporters family	0178	16
CRYSTALLIN BETAGAMMA	Crystallin β and γ 'Greek Key' motif	0184	17
UCH 22	Ubiquitin carboxyl-terminal hydrolases family 2	0190	18
ZINC FINGER C2H2	Zinc finger, C2H2 type	0192	19
TOPOISOMERASE II	DNA topoisomerase II	0192	20
HIS ACID PHOSPHAT	Histidine acid phosphatases	0197	21
G BETA REPEATS	β-transducin family Trp-Asp repeats	0199	22

and their sequences were determined.[1] By ORF analysis, each clone was found to contain a distinct ORF. The ORFs and the first ATG codon are shown in Fig. 1 by solid boxes and open triangles, respectively. In-frame termination codons upstream of the first ATG codon were identified in 20 clones, suggesting that at least 50% of the clones analyzed harbor the complete coding region.

The results of computer analysis with the GCG software package[5] are shown in Tables 1 and 2 and also in the figure in the Supplement section. Sequence features noted are summarized as follows.

1. Sequences of 20 genes were unrelated to any reported genes, except for EST sequences in the database files. The remaining 20 genes carried sequences with some similarities to known genes (Table 1). Among the genes in the latter category, KIAA0167 contained a Zn-finger motif with significant structural similarity to that of the yeast transcription factor *GCS1*[6] (indicated by solid line in Fig. 2A). In ad-

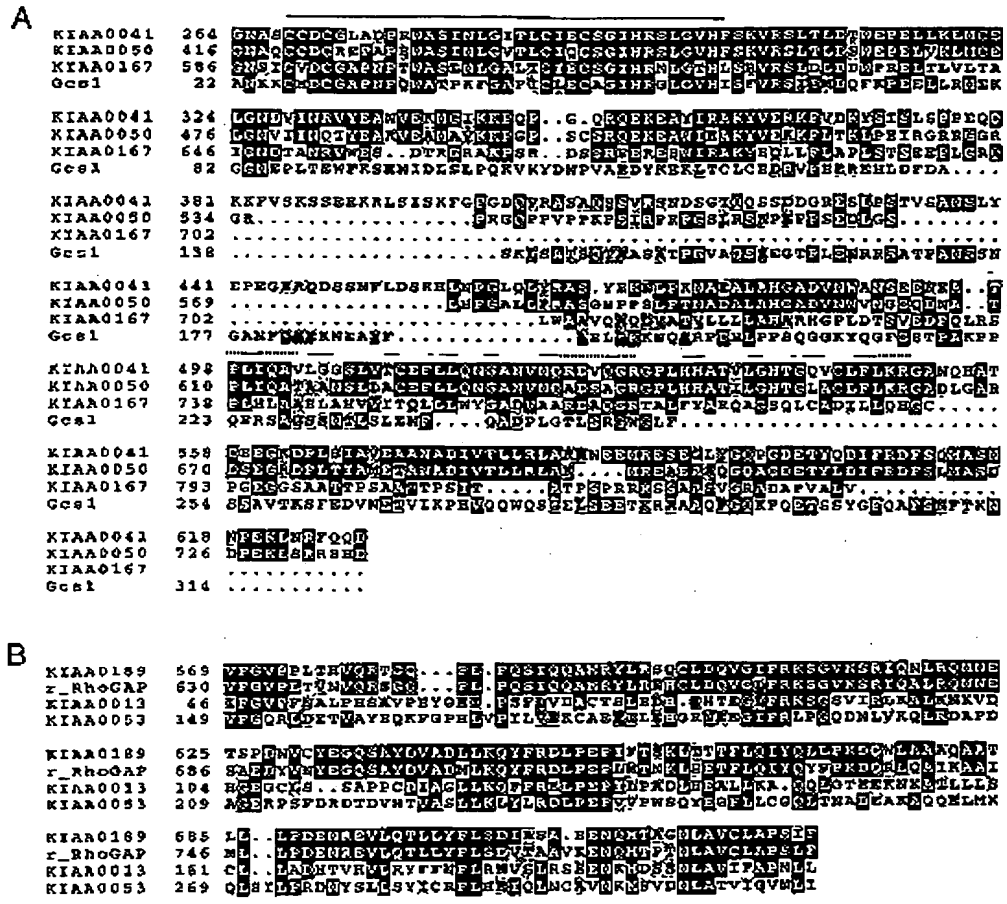


Fig. 2. Sequence comparison in the region spanning the Zn-finger motif (solid line) and ankyrin repeat-like sequence (dotted line) in the KIAA0167 gene family and the *GCS1* gene (A) and that spanning the domain sequence of the RhoGAP and three KIAA clones (B). Identical and similar amino acids are indicated by black and grey background, respectively. Numerals indicate the number of amino acid residues from the start codon.

dition, an ankyrin repeat-like sequence[7] was identified in the adjacent region (indicated by dotted line in Fig. 2A). We noted that two previously reported cDNAs, KIAA0041 and KIAA0050, also carry similar sequences in the corresponding regions. Particular conservation of the Zn-finger motif and ankyrin repeat-like sequence suggests that these genes constitute a novel gene family related to the *GCS1* gene. We tentatively named it the KIAA0167 family. As another gene with similarity to known genes, we noted that KIAA0189 contains the domain sequence of the RhoGAP gene family[8] (GAP: GTPase-activating proteins) (Fig. 2B). Since the sequence conservation between the two genes is quite high, it is likely that KIAA0189 is a member of this family. We also noted that 2 previously reported genes, KIAA0013 and KIAA0053, carry sequences showing weak similarity to the GAP domain (Fig. 2B).

2. Protein motifs that matched those in the PROSITE motif database were found in 11 genes (Table 2).

3. Significant transmembrane domains were identified in 13 genes, 5 of which harbored multiple hydrophobic regions, as judged by the methods of Engelman et al.[9] and of Kyte and Doolittle.[10]

4. Two genes harbored stretches of CAG (Gln) repeats, which were often correlated with genetic disorders:[11] CAG occurred 23 times within a 42 triplet stretch in KIAA0181 and 62 times within a 100 triplet stretch in KIAA0192 (Fig. 3A and B).

5. Another novel repeat composed of alternating Arg

Table 3 Summary of cDNA sequence data and expression patterns of cloned genes in human tissues and cell lines.

Gene number (KIAA)	Total length of cDNA (bp)[a]	Amino acid residues	Expression[b] KG-1	HeLa	He	Br	Pl	Lu	Li	Sk.m	Ki	Pa	Sp	Th	Pr	Te	Ov	Sm.i	Co	Pc.b	Chromosomal location	Accession[c] number
0161[d,f]	5,559	292	+	−	+	±	+	+	−	+	+	+	+	−	−	±	+	±	−	=	2	D79983
0162[e]	5 876	1.726	+	+	+	±	+	±	+	++	±	+	±	±	+	+	+	+	+	±	17	D79984
0163[d]	4,436	550	+	+	+	±	+	+	+	+	+	±	±	±	+	+	+	+	+	±	22	D79985
0164[e]	5 538	920	+	+	+	±	+	±	+	++	±	±	±	+	+	+	±+	+	+	±	6	D79986
0165[d,e]	6 662	1.795	+	+	−	−	−	−	−	−	−	−	−	+	−	++	−	+	+	−	8	D79987
0166[d]	6 942	2 209	+	+	+	±	+	±	±	±	±	±	±	+	±	++	±	+	+	+	17	D79988
0167[e,f]	3 950	836	+	−	−	+++	−	±	±	+	−	−	±	+	−	−	−	±	−	+	12	D79989
0168	5 426	326	+	−	−	±±	±	++	+	+	±	−	±	+	±	+	±	+	−	±+	20	D79990
0169[d]	5 656	1.745	+	+	+	+	±	+	+	+	±	+	+	+	−	+++	±	+	±	+	9	D79991
0170[e]	6,940	2 089	+	+	+	+	±	++	±	++	±+	+	±	+	±	+++	+	+	±	+	6	D79992
0171[e]	3 336	625	+	+	+	+	±	+	+	++	±	±	±	+	±	+	±	+	±	+	5	D79993
0172[e]	1 792	1.307	+	−	++	±	+	+	+	+	±	±	±	+	±+	+	++	+	+	−	9 20	D79994
0173[e,f]	4,831	1,199	+	+	+	±	+	+	±	+	±	+	±	+	±	++	±	+	+	+	2	D79995
0174	2.348	364	+	+	+	+	+	±	+	+	++	±	±	±	±	±	+	+	+	+	19	D79996
0175[d,e,f]	2,470	651	+	+	−	−	+	−	−	−	±	−	−	±	−	+	±	±	+	−	9	D79997
0176	3.635	265	±	±	+	+	+	±	±	±	+	±	+	+	±	±	±	+	+	+	17	D79998
0177[d,e]	5.083	1.631	±	±	++	+	+	++	±	±	++	±	+	±	±	±	+	+	+	++	13,22	D79999
0178[e,f]	5.803	1.225	+	±	+	±	±	±±	±	±	+	±	+	+	+	±	±	±	+	+	X	D80000
0179[e]	4.994	762	±	±	+	+	+	±	−	±	+	±	+	±	±	±	±	±	+	+	21	D80001
0180	5,009	343	±	±	++	±	±	±	−	±	+	±	+	+	+	++	+	±	±	+	20	D80002
0181	6,504	2,005	±	+	±	+	+	++	−	+	+	+	±	+	+	±	+	+	±	±	20	D80003
0182	7,133	1,157	±	+	±	±	±	±	±	+	+	+	±	+	+	−	+	+	±	+	16	D80004
0183	4,905	1,062	++	++	++	+	++	++	++	++	++	±−	±±	++	++	±	++	±	+	++	9	D80005
0184[f]	4,639	863	±	+	±	+	+	±	±	+	±	±	−	+	+	±	+	±	+	±	2	D80006
0185[e]	5,823	1,884	+	+	±	±	+	±	±	±	±	±	−	+	−	±±	+	±	+	±	10	D80007
0186	3,248	196	±	+	−	−	−	−	−	−	−	−		±	−	±	−	±	±	−	20	D80008
0187[f]	4,181	1,282	+	+	±±	+	+	+	+	++	−	±	±	±	+	±−	+	−	+	±	10	D80009
0188[e]	5,307	899	±	+	+	+	+	+	+	+++	±	+	±	±	±	±±	+	±	+	±	2	D80010
0189[e]	4,824	1,023	+	−	+	−	+	+	+	+	±	±	±	±	+	+	+	+	±	−	X	D80011
0190[e,f]	3,280	839	+	±	+	±	+	+	±	±±	±	±	±	±	+	++	±	+	±	+	16,22	D80012
0191[f]	5 203	1 516	+	±	+	±	+	+	±	±±	±	±	±	±	+	++	−	+	±	±	1	D83776
0192[d,e,f]	6.604	2.124	+	±	+	±	+	±	±	±	+	±	±	±	±	±	±	±	±	±	X	D83783

Table 3. Continued.

Gene number (KIAA)	Total length of cDNA (bp)[a]	Amino acid residues	Expression[b]																		Chromosomal location	Accession[c] number
			KG-1	HeLa	He	Br	Pl	Lu	Li	Sk.m	Ki	Pa	Sp	Th	Pr	Te	Ov	Sm.i	Co	Pe.b		
0193	5,076	346	+	±	+	+++	+	+	±	±	+	+	+	+	+	+	+	+	±	+	7	D83777
0194	5,245	1,435	+	±	+	±	±	±	±	±	±	+	+	+	+	+	+	+	+	+	4.5	D83778
0195[d]	5,022	1,356	+	±	+	+	+	+	±	++	±	±	+	+	+	++	+	+	±	+	17	D83779
0196[d]	4,103	1,159	+	±	+	+	±	+	+	++	±	±	±	+	+	+	+	+	+	+	8	D83780
0197[d,f]	4,814	1,314	+	±	+	±	±	+	+	+	±	+	±	+	+	+	+	+	+	+	11	D83781
0198[d,e]	5,638	562	+	+	+	−	−	−	−	+	−	±	±	±	−	+	+	+	+	+	20	D83784
0199[d,e,f]	3,976	1,277	+	+	+	±	±	+	+	+	++	+	±	±	+	+	++	+	+	+	3	D83782
0200[e]	5,713	1,016	+	+	+	±	+	+	+	+	±	+	+	±	+	+	+	±	+	+	5	D83785

n.d., not determined; He, heart; Br, brain; Pl, placenta; Lu, lung; Li, liver; Sk.m, skeletal muscle; Ki, kidney; Pa, pancreas; Sp, spleen; Th, thymus; Pr, prostate; Te, testis, Ov, ovary; Sm.i, small intestine; Co, colon; Pe.b, peripheral blood leukocytes. [a] Values excluding poly(A) sequences. [b] Expression of mRNA in indicated cells and human tissues (Clontech, USA) was examined by northern hybridization, and the strength of the positive signals are indicated (±, +, ++, +++). [c] Accession number of GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases. [d] Putative transmembrane domains were contained (see Supplemental pages). [e] Similarities to known genes were identified (see Table 1 and Supplemental pages). [f] Protein motifs were recognized (see Table 2 and Supplemental pages).

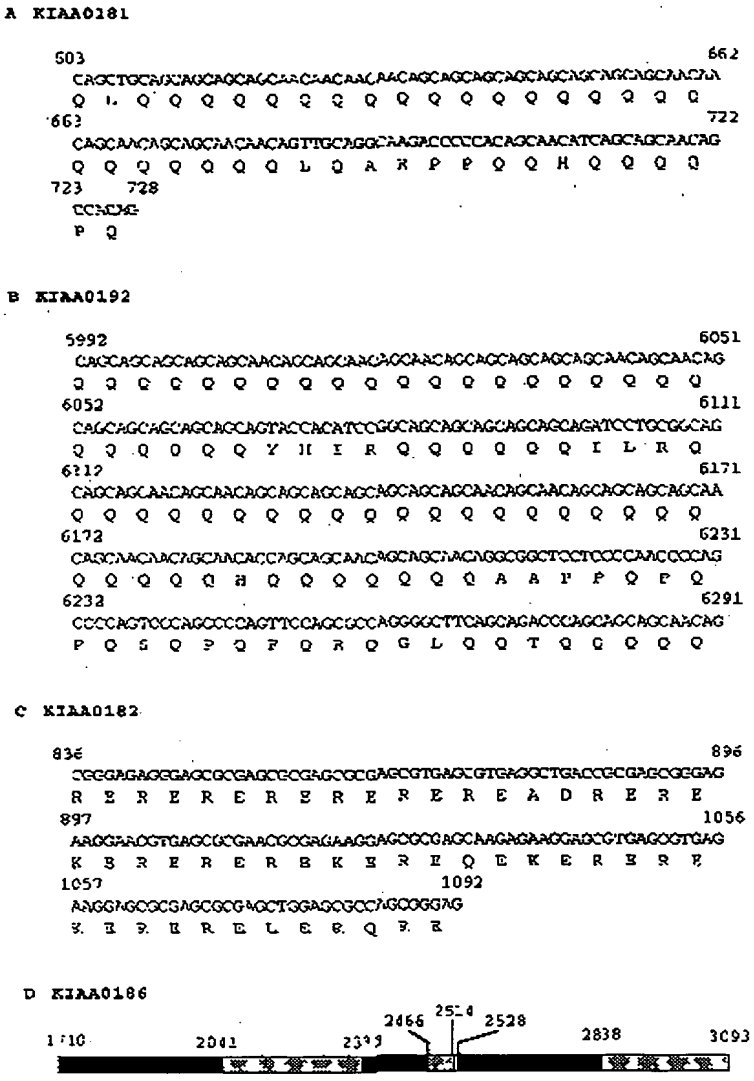


Fig. 3. Typical repeats observed in cDNA clones. A and B, CAG repeats in KIAA0181 and KIAA0192: C, RE repeats identified in KIAA0182: D, a complex array of Alu repeats (solid boxes) and partial LINE sequence (dotted boxes) found in the 3′-UTR of KIAA0186. In A, B and C, translated amino acids are indicated below the DNA sequences. Numerals above the sequences are the nucleotide positions in each clone.

and Glu (RE repeat) was observed in KIAA0182 (Fig. 3C). Alternating Arg-Asp tracts have been found in many RNA-binding proteins[12] as a characteristic sequence. The presence of the RE repeat with structural similarity to the RD repeat strongly suggests that KIAA0182 exhibits an RNA-binding activity.

6. Alu sequences were identified in the 3′-untranslated region (UTR) of 6 genes, including KIAA0186 which retained a complex array of repetitive sequences. In the 3′-UTR of this gene, both Alu repeats and partial LINE sequences reiterated 3 times, respectively (Fig. 3D).

3.2. *Expression profiles in tissues*

The expression profiles of the sequenced genes were examined with 16 different human tissues and 2 cell lines including the KG-1 cell as a positive control. The results are summarized in Table 3. Thirty genes were expressed ubiquitously in all the cells and tissues examined. The

remaining 10 genes apparently showed different expression profiles among the cells and tissues examined. Although the spectra were different for each gene, most of them seemed to belong to the class of genes of which the expression was specifically suppressed in certain tissues. In contrast, it was significant that 4 genes, KIAA0165, 0167, 0175 and 0186 were only expressed in a few specific tissues.

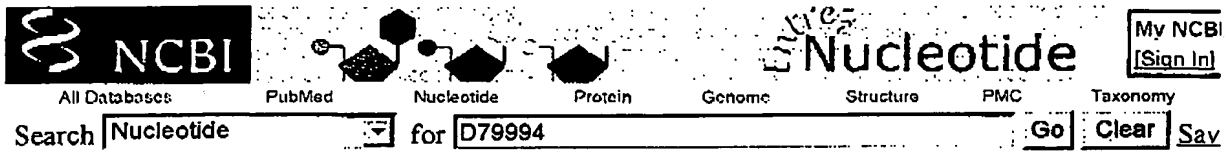
As seen in the previously reported genes (Fig. 2 in ref. 2, 3), multiple but discrete bands were recognized in 30 clones, possibly due to either alternative splicing, alternative termination, or initiation of transcription at different sites.

The chromosomal location of these genes have been determined using a panel of human-rodent hybrid cell lines (see Table 3).

Acknowledgments: This project was supported by grants from the Kazusa DNA Research Institute Foundation. We thank Dr. M. Takanami for his support of this work.

References

1. Nomura, N., Miyajima, N., Sazuka, T. et al. 1994, Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1, *DNA Res.*, **1**, 27–35, Supplement; **1**, 47–56.
2. Nomura, N., Nagase, T., Miyajima, N. et al. 1994, Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1, *DNA Res.*, **1**, 223–229, Supplement; **1**, 251–262.
3. Nagase, T., Miyajima, N., Tanaka, A. et al. 1995, Prediction of the coding sequences of unidentified human genes. III. The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by analysis of cDNA clones from human cell line KG-1, *DNA Res.*, **2**, 37–43, Supplement; **2**, 51–59.
4. Nagase, T., Seki, N., Tanaka, A. et al. 1995, Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1, *DNA Res.*, **2**, 167–174, Supplement; **2**, 199–210.
5. Devereux, J., Haeberli, P., and Smithies, O. 1984, A comprehensive set of sequence analysis programs for the VAX, *Nucleic Acids Res.*, **12**, 387–395.
6. Ireland, L. S., Johnston, G. C., Drebot, M. A. et al. 1994, A member of a novel family of yeast Zn-finger proteins mediates the transition from stationary phase to cell proliferation, *EMBO J.*, **13**, 3812–3821.
7. Lambert, S., Yu, H., Prchal, J. T. et al. 1990, cDNA sequence for human erythrocyte ankyrin, *Proc. Natl. Acad. Sci. USA.*, **87**, 1730–1734.
8. Homma, Y. and Emori, Y. 1995, A dual functional signal mediator showing RhoGAP and phospholipase C-δ stimulating activities, *EMBO J.*, **14**, 286–291.
9. Engelman, D. M., Steize, T. A., and Goldman, A. 1986, Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins, *Annu. Rev. Biophys. Biophys. Chem.*, **15**, 321–353.
10. Kyte, J. and Doolittle, R. F. 1982, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.*, **157**, 105–132.
11. Mandel, J.-L. 1994, Trinucleotide diseases on the rise, *Nature genetics*, **7**, 453–455.
12. Surowy, C. S., Hoganson, G., Gosink, J., Strunk, K., and Spritz, R. A. 1990, The human RD protein is closely related to nuclear RNA-binding proteins and has been highly conserved, *Gene*, **90**, 299–302.
13. Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P., and Adams, J. M. 1991, Novel zinc finger gene implicated as *myc* collaborator by retrovirally accelerated lymphomagenesis in Eμ-*myc* transgenic mice, *Cell*, **65**, 753–763.
14. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. 1982, Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, *EMBO J.*, **1**, 945–951.
15. Hunter, T. 1991, Protein Kinase Classification, In: Hunter, T., Sefton, B. M. (eds) Methods in Enzymology, Vol. 200. Academic Press, New York, pp. 3–37.
16. Higgins, C. F., Hiles, I. D., Salmond, G. P. C. et al. 1986, A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria, *Nature*, **323**, 448–450.
17. Wistow, G. 1993, Lens crystallins: gene recruitment and evolutionary dynamism, *TIBS*, **18**, 301–306.
18. Papa, F. R. and Hochstrasser, M., 1993, The yeast *DOA4* gene encodes a deubiquitinating enzyme related to a product of the human *tre-2* oncogene, *Nature*, **366**, 313–319.
19. Rosenfeld, R. and Margalit, H. 1993, Zinc fingers: conserved properties that can distinguish between spurious and actual DNA- binding motifs, *J. Biomol. Struct. Dyn.*, **11**, 557–570.
20. Wyckoff, E., Natalie, D., Nolan, J. M., Lee, M., and Hsieh, T., 1989, Structure of the *Drosophila* DNA Topoisomerase II gene: Nucleotide sequence and homology among Topoisomerase II, *J. Mol. Biol.*, **205**, 1–13.
21. Van Etten, R. L., Davidson, R., Stevis, P. E., MacArthur, H., and Moore, D. L., 1991, Covalent structure, disulfide bonding, and identification of reactive surface and active site residues of human prostatic acid phosphatase, *J. Biol. Chem.*, **266**, 2313–2319.
22. Gilman, A. G. 1987, G proteins: Transducers of receptor-generated signals, *Annu. Rev. Biochem.*, **56**, 615–649.



About Entrez

Entrez Nucleotide
Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources
BLAST

Reference sequence project

Search for Genes

Submit to GenBank

Search for full length cDNAs

Show only records from: CoreNucleotide (1), EST (0), GSS (0). [What's this?]

☐ 1: D79994 Reports Links
Homo sapiens KIAA0172 mRNA, complete cds
gi|58257638|dbj|D79994.2|[58257638]

1: D79994. Reports Homo sapiens KIAA...[gi:58257638] Links

Comment Features Sequence

```
LOCUS       D79994                  5635 bp    mRNA    linear   PRI 28-JAN-2005
DEFINITION  Homo sapiens KIAA0172 mRNA, complete cds.
ACCESSION   D79994
VERSION     D79994.2  GI:58257638
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
            Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
            Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
            Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
            chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5635)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
            Kazusa-kamatari 2-6-7, Kisarazu, Chiba, 292-0818, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:81-438-52-3913)
COMMENT     On Jan 27, 2005 this sequence version replaced gi:1136403.
FEATURES             Location/Qualifiers
     source          1..5635
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02512s1"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1-1718 was derived
                     from hh14360 and 1719-5635 was derived from ha2512)."
     gene            1..5635
                     /gene="KIAA0172"
```

```
CDS             <667..4767
                /gene="KIAA0172"
                /note="Start codon is not identified
                similar to ankyrin of Chromatium vinosum."
                /codon_start=1
                /protein_id="BAA11489.2"
                /db_xref="GI:58257639"
                /translation="LLTPFWISHWTQASMAHTTKVNGSASGKAGDILSGDQDKEQKDP
                YFVETPYGYQLDLDFLKYVDDIQKGNTIKRLNIQKRRKPSVPCPEPRTTSGQQGIWTS
                TESLSSSNSDDNKQCPNFLIARSQVTSTPISKPPPPLETSLPFLTIPENRQLPPPSPQ
                LPKHNLHVTKTLMETRRRLEQERATMQMTPGEFRRPRLASFGGMGTTSSLPSFVGSGN
                HNPAKHQLQNGYQGNGDYGSYAPAAPTTSSMGSSIRHSPLSSGISTPVTNVSPMHLQH
                IREQMAIALKRLKELEEQVRTIPVLQVKISVLQEEKRQLVSQLKNQRAASQINVCGVR
                KRSYSAGNASQLEQLSRARRSGGELYIDYEEEEMETVEQSTQRIKEFRQLTADMQALE
                QKIQDSSCEASSELRENGECRSVAVGAEENMNDIVVYHRGSRSCKDAAVGTLVEMRNC
                GVSVTEAMLGVMTEADKEIELQQQTIEALKEKIYRLEVQLRETTHDREMTKLKQELQA
                AGSRKKVDKATMAQPLVFSKVVEAVVQTRDQMVGSHMDLVDTCVGTSVETNSVGISCQ
                PECKNKVVGPELPMNWWIVKERVEMHDRCAGRSVEMCDKSVSVEVSVCETGSNTEESV
                NDLTLLKTNLNLKEVRSIGCGDCSVDVTVCSPKECASRGVNTEAVSQVEAAVMAVPRT
                ADQDTSTDLEQVHQFTNTETATLIESCTNTCLSTLDKQTSTQTVETRTVAVGEGRVKD
                INSSTKTRSIGVGTLLSGHSGFDRPSAVKTKESGVGQININDNYLVGLKMRTIACGPP
                QLTVGLTASRRSVGVGDDPVGESLENPQPQAPLGMMTGLDHYIERIQKLLAEQQTLLA
                ENYSELAEAFGEPHSQMGSLNSQLISTLSSINSVMKSASTEELRNPDFQKTSLGKITG
                NYLGYTCKCGGLQSGSPLSSQTSQPEQEVGTSEGKPISSLDAFPTQEGTLSPVNLTDD
                QIAAGLYACTNNESTLKSIMKKKDGNKDSNGAKKNLQFVGINGGYETTSSDDSSSDES
                SSSESDDECDVIEYPLEEEEEEEEDEDTRGMAEGHHAVNIEGLKSARVEDEMQVQECEP
                EKVEIRERYELSEKMLSACNLLKNTINDPKALTSKDMRFCLNTLQHEWFRVSSQKSAI
                PAMVGDYIAAFEAISPDVLRYVINLADGNGNTALHYSVSHSNFEIVKLLLDADVCNVD
                HQNKAGYTPIMLAALAAVEAEKDMRIVEELFGCGDVNAKASQAGQTALMLAVSHGRID
                MVKGLLACGADVNIQDDEGSTALMCASEHGHVEIVKLLLAQPGCNGHLEDNDGSTALS
                IALEAGHKDIAVLLYAHVNFAKAQSPGTPRLGRKTSPGPTHRGSFD"
ORIGIN
        1 gcggccgaag aaaaacagga caaagaaacc tgtgtgaatt ttcagctttt caccttctct
       61 gattttattt cttcctcact cttcctttga cagtcctcgc agtccgcctg agagagtaga
      121 gaagaccccc tcccagaacc ttcctgtaag gtctccccgc tgacttcctg tagtggatgt
      181 gactgtgtgc caggtgcctg ccgaagaccc ctgtcactga ctgtgccctt tgggaaagag
      241 tcaacaatgg cctcctccta tggctgactt cgttcatgtt caggctctcc agcttgctag
      301 tgagaattgc caaggattgg ttcatggcag gatagaacta aactgataga tgaaagctcc
      361 acagtgcttc aagacggggt cccgttcatg ttaagcagtt ttccttcctt caaaataaaa
      421 aactagcagt ccttagggag gacagttttt tccttccttt tcctattctg gctccaacag
      481 ctgtctcaac acagcccgag atggggagca gcctggcttc caccggcaat agctgtattg
      541 tgggagtgtg aaagagaagc caccttttcc tccctctgcc caagccacct ggccccttg
      601 tcctttctcc tcctcgtcct ctgaggttga atgcctttga gaacttgatg cataaaattt
      661 gcatgactcc tcactccttt ctggatctcc cattggactc aagccagcat ggctcacacc
      721 acaaaggtta acggcagtgc ctcaggaaaa gcaggtgata ttctcagtgg agaccaggac
      781 aaggaacaga aagaccctta ctttgtggag acccctatg gttatcaact agacttagat
      841 ttcctcaaat atgtggatga catacagaag ggaaatacca tcaaaagact gaacatccag
      901 aagaggcgga agccgtccgt gccatgccca gaacccagga ccacatctgg tcagcaaggt
      961 atatggactt ccactgaatc cctctcatcc tccaacagtg atgacaacaa gcagtgcccc
     1021 aacttcctca tagccagaag tcaagttaca tcaactccaa tctcaaagcc acctcccct
     1081 ctggagacct cactcccttt tcttaccatc ccagaaaatc gacagctgcc acctccctca
     1141 ccacaactcc caaagcataa ccttcatgtc accaagacac tgatggagac ccggagaaga
     1201 ctggaacagg agagagccac catgcagatg acaccgggtg agttcagaag gcccaggctg
     1261 gccagttttg gaggcatggg caccacaagc tccctccctt cttttgtggg ttctggaaac
     1321 cacaatcctg ccaagcacca gcttcagaat ggataccaag gtaatgggga ttatggtagc
     1381 tatgccccag ctgctcccac cacttcctcc atggggagct ccatccgcca cagccccctg
     1441 agctcaggga tctccacccc agtgaccaac gtgagcccca tgcacctgca gcacatccgc
     1501 gagcagatgg ccattgctct gaaacgcctg aaggagctgg aggagcaggt gcgaaccatc
     1561 cctgtgctcc aggtaaagat ctctgtcttg caagaagaga aaaggcagtt ggtctcacag
     1621 ctgaaaaacc aaagggctgc atcccagatc aatgtctgtg gtgtgaggaa gcggtcctat
     1681 agtgcgggga acgcctccca gctggaacag ctctcccggg cccgaagaag tggcggggaa
     1741 ttatacattg actatgagga ggaagaaatg gagaccgtag aacagagcac gcagaggata
     1801 aaggagttcc ggcaacttac agcagacatg caagccctgg agcagaagat ccaggacagc
     1861 agctgtgagg cctcctcaga gctcagggag aatggagagt gccggtctgt ggctgtgggt
```

```
1921 gccgaggaga acatgaacga catcgtcgtg taccacagag gctccaggtc ctgtaaggat
1981 gcagctgtag ggacacttgt tgagatgaga aattgtgggg tcagcgtgac agaggccatg
2041 cttggagtga tgactgaagc tgacaaagaa attgagctcc aacagcagac catagaagcc
2101 ttgaaggaaa agatctatcg cctagaagta cagcttagag aaaccaccca tgaccgggag
2161 atgactaaac tgaaacaaga gctgcaggct gctggatcga ggaaaaaggt tgacaaagcc
2221 acgatggccc agccgcttgt tttcagtaag gtggtggagg cagtggtgca gaccagagac
2281 caaatggtcg gcagtcacat ggacctggtg gacacgtgtg ttgggacctc cgtggaaaca
2341 aacagtgtag gcatctcctg ccagcctgaa tgtaagaata aagtcgtagg gcctgagctg
2401 cctatgaatt ggtggattgt taaggagagg gtggaaatgc atgaccgatg tgctgggagg
2461 tctgtggaaa tgtgtgacaa gagtgtgagt gtggaagtca gcgtctgcga aacaggcagc
2521 aacacagagg agtctgtgaa cgacctcaca ctcctcaaga caaacttgaa tctcaaagaa
2581 gtgcggtcta tcggttgtgg agattgttct gttgacgtga ccgtctgctc tccaaaggag
2641 tgcgcctccc ggggcgtgaa cactgaggct gttagccagg tggaagctgc cgtcatggca
2701 gtgcctcgta ctgcagacca ggacactagc acagatttgg aacaggtgca ccagttcacc
2761 aacaccgaga cggccaccct catagagtcc tgcaccaaca cttgtctaag cactttggac
2821 aagcagacca gcacccagac tgtggagacg cggacagtag ctgtaggaga aggccgtgtc
2881 aaggacatca actcctccac caagacgcgg tccattggtg ttggaacgtt gctttctggc
2941 cattctgggt ttgacaggcc atcagctgtg aagaccaaag agtcaggtgt ggggcagata
3001 aatattaacg acaactatct ggttggtctc aaaatgagga ctatagcttg tgggccacca
3061 cagttgactg tggggctgac agccagcaga aggagcgtgg gggttgggga tgaccctgta
3121 ggggaatctc tggagaaccc ccagcctcaa gctccacttg gaatgatgac tggcctggat
3181 cactacattg agcgtatcca gaagctgctg gcagaacagc agacactgct ggctgagaac
3241 tacagtgaac tggcagaagc tttcggggaa cctcactcac agatgggctc cctcaactct
3301 cagctcatca gcaccctgtc gtctatcaac tctgtcatga aatctgcaag cactgaagag
3361 ctgaggaacc ctgacttcca gaaaaccagt ctgggtaaaa tcacaggcaa ttatttggga
3421 tatacctgta agtgtggggg ccttcagtca ggaagtccct taagctccca gacatcccag
3481 cctgagcaag aagtggggac ctcagaagga aagccaatca gcagcctgga tgccttcccc
3541 actcaggaag gtacgctgtc tccagtgaac ctgacagacg accagatcgc cgctggcctc
3601 tatgcatgta caaacaatga aagtacactg aagtccatca tgaagaagaa agatggtaac
3661 aaagattcaa atggcgcaaa aaagaatctt cagtttgttg gcattaatgg agggtatgaa
3721 acaacttcaa gtgatgattc cagctcagat gaaagctctt cttccgagtc agatgacgag
3781 tgtgatgtca ttgagtatcc tcttgaagaa gaggaggagg aggaggatga agacactcgg
3841 ggaatggcag aagggcacca tgcagttaat attgaaggtt tgaagtctgc cagggtggaa
3901 gatgaaatgc aggttcaaga atgtgaacct gagaaggtgg aaatcagaga gaggtatgaa
3961 ttaagtgaaa agatgttgtc tgcatgcaac ttactgaaaa atactataaa tgaccccaaa
4021 gctttgacca gcaaagatat gaggttctgt ctgaacaccc tccagcacga gtggttccgc
4081 gtgtccagtc agaagtcagc cattccagcc atggtggggg actacatagc tgcttttgag
4141 gccatttccc cagatgtcct ccgctatgtc atcaacttgg cagacggcaa cggcaacaca
4201 gccctccatt acagcgtgtc ccactccaac ttcgagattg tgaagctgct gttagatgcc
4261 gatgtgtgta atgtggatca ccagaacaag gcaggctaca cccccatcat gttggcggcc
4321 ctcgccgctg tggaagcaga gaaggacatg cggattgtgg aagaactctt tggctgtggg
4381 gatgtgaatg ccaaagctag tcaggcggga cagacggccc tcatgctggc ggtcagtcac
4441 ggacggatag acatggtgaa gggccttctg gcctgtgggg ctgatgtcaa catccaggat
4501 gacgagggct ccacggccct catgtgtgcc agcgagcacg ggcacgtgga gattgtcaag
4561 ctgctgctgg cccagcccgg ctgcaacggt cacctagagg acaacgatgg cagcactgcg
4621 ctctcaatcg ccctggaagc aggacacaag gacatcgctg ttcttctgta tgcccatgtc
4681 aactttgcaa aagcccagtc tccgggcacc cctaggcttg gaaggaagac gtctcctggc
4741 cccacccacc gaggttcatt tgattgattg tatgcaaata gccctttatt tacatgccac
4801 tattaagctg ctaattgttc ctgttggggt gacagatact gaatgtatac gtattgtgcc
4861 tgagctcacc agcaaacaga agcatcaagc ccaggggtaa aggctgaagc tttcacagtg
4921 cagagactgc tagcctgggc acacgcacct cctttctggc cgtcttctgt gtagggcaca
4981 ctttaaccca gtctctgttg ctgttgagtc tctgctccgt tttgtacagt cacagggaat
5041 tctgatctga aggggcacct tctgttcact cccacaaagt ggtgtctggt tctcactgag
5101 acgttttaag atttttccac aaatatttat atgtactaaa tgtggaacca ttagaaagtt
5161 cttccaaaat ctcattccag catagttttg gatttttctt ttgtcttatt ttaaaataag
5221 gaagtcgaga tgactttgat cattggtaac ttgggcctgg gccagacaaa gtataaaact
5281 tacaaaagaa tattctcatt tggtcttaac taggtagatg taatatatga ctttttataa
5341 aaagggtatc tatatgaact tgacacagta ttttcagctt ttgtattcca tactaaagcc
5401 atgaagaact acacgtaaca tcatcatttg tattaattgc acaactccaa tgctaaaggt
5461 tggattgtgt tagaggaatc ggctctgtat ttgcctctag agaaacacag tgttctcttt
5521 gtatttatgg attccttttt accgtgtcac atttactttg gtcctctatg tatttaaatg
5581 tttgaagtgc cttagactct tgccatattt tcaaaataaa attccattaa gctct
//
```